Pole-Zero Fitting for Transfer Function of Hearing-Aid Receiver: Evidence-Based Review

Woojae Han[1] and Noori Kim[2]

[1]Laboratory of Hearing and Technology, Division of Speech Pathology and Audiology, Research Institute of Audiology and Speech Pathology, College of Natural Sciences, Hallym University, Chuncheon, Korea
[2]Department of Electrical and Computer Engineering, DigiPen Institute of Technology Singapore, Singapore

The hearing-aid transducer is ubiquitous in the hearing-aid industry. For example, the balanced armature receiver (BAR), first invented by A.G. Bell, has been used in all telephone earphones because it has the highest output and best frequency response. Nevertheless, previous electro-mechanical studies on these miniature speakers are quite primitive, given the price of the transducers. Thus, more detailed analysis is critically important for the field of hearing science. This review study was motivated by Hunt's parameter calibration (1954), a widely used commercial hearing-aid receiver (ED series, manufactured by Knowles Electronics, Inc.). In the body of the study, the transfer function of the BAR system (i.e., pressure over voltage) was calculated from Hunt's parameters, solely from the electrical terminals of the device. The computed transfer function was then further investigated by comparing to the pole-zero fitting method using the methods of Gustavsen and Semlyen (1999) and Prony (1975). Based on our short experiment, the better fitting result was achieved with Gustavsen and Semlyen's method. By decomposing results of the transfer function fitting into all-pass and minimum-phase parts, the system was confirmed as a delay system. We conclude that the BAR system is linear, time-invariant, stable, and causal while providing an evidence-based understanding of the hearing-aid receiver system. **J Audiol Otol 2018;22(3):111-119**

KEY WORDS: Pole-zero fitting · Hearing-aid · Receiver · Transfer function.

Introduction

In general, a hearing-aid consists of three parts: a microphone (picks up sound), an amplifier (transforms sound into different frequencies, filters noise, and selectively amplifies each frequency region based on the difference in individuals with hearing loss via multi-band compression), and a receiver (sends the processed signal from the amplifier into the ear). A better understanding of each component of the hearing-aid can improve its sound quality. In this review paper, we analyze the hearing-aid receiver, which is one of the most important, expensive, and complex components of the hearing-aid; specifically, we study the balanced armature receiver (BAR), which converts an electrical signal (current) into acoustical pressure (or force in the case of an electro-mechanical system). In particular, while introducing a unique analysis procedure for the BAR system, the entire study can be summarized into three parts: 1) calculating the system's transfer function from Hunt's parameters, 2) pole-zero fitting of the transfer function, and 3) decomposition of the transfer function into all-pass and minimum-phase parts. This concept is inspired by Robinson, et al. [1], who approximated the length of the ear canal using the human ear's reflectance data.

Beyond Hunt's classic study [2], several papers have implemented anti-reciprocity in their electro-magnetic transducer model. In 2013, Kim and Allen [3] suggested a two-port anti-reciprocal network model of the BAR having a semi-inductor, a gyrator (two poorly understood elements of special interest in the electro-magnetic transducer), and a pure delay. Their two-port impedance matrix is based on the classical expression of either Wegel [4] or Hunt [2] to transform electrical to acoustic segments,

$$\begin{bmatrix} E(\omega) \\ P(\omega) \end{bmatrix} = \begin{bmatrix} Z_e(s) & T_{ea}(s) \\ T_{ae}(s) & Z_a(s) \end{bmatrix} = \begin{bmatrix} I(\omega) \\ U(\omega) \end{bmatrix}, \qquad (1)$$

where E is the voltage, I is the current, P is the pressure, and U is the volume velocity; the directions of I and U are defined as going into the network.

Note that $s = \sigma + j\omega$, and

$$Z_e(s) = \frac{\Phi(\omega)}{I(\omega)} \quad \text{when } U(\omega)=0, \qquad (2)$$

$$T_{ea}(s) = -T_a(s) = \frac{E(\omega)}{U(\omega)} \quad \text{when } I(\omega)=0, \qquad (3)$$

$$T_{ae}(s) = T_a(s) = \frac{P(\omega)}{I(\omega)} \quad \text{when } U(\omega)=0, \qquad (4)$$

$$Z_a(s) = \frac{P(\omega)}{U(\omega)} \quad \text{when } I(\omega)=0. \qquad (5)$$

The transduction functions, $T_{ea}(s)$ and $T_{ae}(s)$, show anti-reciprocal characteristics of the system. For example, under the DC status (ignoring acoustic components), the system's electro-magnetic coupling coefficient is a constant, $-T_{ea} = T_{ae} = -T_a = B_0 l$ where B_0 and l are the DC magnetic field and length of wire, respectively. The $T_{ea}(s)$ and $T_{ae}(s)$ have the same magnitudes but opposite signs.

Along with Eq. 1, the two-port electro-mechanic transducer equation can alternatively be represented in ABCD (a.k.a. transmission matrix) form, as given by

$$\begin{bmatrix} E(\omega) \\ I(\omega) \end{bmatrix} = \begin{bmatrix} A(s) & B(s) \\ C(s) & D(s) \end{bmatrix} \begin{bmatrix} P(\omega) \\ -U(\omega) \end{bmatrix}. \qquad (6)$$

Here A, B, C, and D are functions of s to show that they are causal and complex analytic system variables. The signal variables E, I, P, and U, on the other hand, are functions of ω, indicating they are signals and therefore neither causal nor analytic.

The fundamental difference between the two matrix representations lies in the coupling of acoustic and electric segments. Specifically, the electrical input parameters E and I are on the left side of the network (Eq. 6) and expressed in terms of the acoustical variables, P and U, on the right side of the network, via the four frequency-dependent parameters A, B, C, and D. Conversion between Eq. 1 and Eq. 6 yields the following relationships:

$$\begin{bmatrix} A(s) & B(s) \\ C(s) & D(s) \end{bmatrix} = \frac{1}{T_{ae}(s)} \begin{bmatrix} Z_e(s) & \Delta_Z \\ 1 & Z_a(s) \end{bmatrix}, \qquad (7)$$

where $\Delta_Z = Z_e Z_a - T_{ea} T_{ae}$. Note that if C=0, Z does not exist.

Using a friendly approach, this paper is organized into four sections: First section introduces the theoretical background. The second section includes the pole-zero fitting algorithm and its derivation. In third and fourth sections, we present the experimental methods and their results, respectively. Conclusion is followed by the final section.

Theoretical Background

This section introduces the most critical theories and concepts needed to appreciate this study. These include concepts such as two-port network postulates and transfer functions with equivalent forms.

Two-port network theory

The impedance matrix representation of two-port networks (Eq. 1) is the matrix version of Ohm's law. Like the one-port Thevenin/Norton model, the matrix system can characterize and generalize linear transducers [2,4,5]. Carlin and Giordano [6] summarized two-port networks in terms of six postulates: 1) Linearity (vs. nonlinearity)—A system obeys superposition. 2) Time-invariance (vs. time-variance)—A system does not depend on the time of excitation. 3) Passivity (vs. active)—This addresses the conservation of energy law; a system cannot provide more power than the supplied amount. 4) Causality (vs. non-causality vs. anti-causality)—A system response cannot be affected by a future response. 5) Real-time function (vs. complex-time function)—The system's time response is real. 6) Reciprocity (vs. non-reciprocity vs. anti-reciprocity)—The two transfer impedances (the two off-diagonal components) of the systems impedance matrix must be equal to be a reciprocal system. The anti-reciprocal network swaps the force and the flow, but one variable changes to the opposite direction.

Based on a study of McMillan [7] and Hunt [2], a moving-armature electro-magnetic transducer can be approximately considered to be a linear, time-invariant, passive, causal, real-time, and anti-reciprocal system (the off-diagonal elements if Eq. 1 has opposite signs). Therefore, this system is linear time-invariant (LTI) and capable of being analyzed using Laplace and Fourier transforms. In transform domains, the linear differential equations of the system (in the time domain) become linear and algebraic, so we can compute the algebra to study the LTI system. The linearity is the most important concept in the system theory. Many practical engineering systems can be designed as linear. This concept allows for easier and more convenient methods of system analysis and modeling. Time-invariance, as discussed above, means that if the same input is applied at different times, the output produced will

be identical in shape and size but will be shifted in time.

Transfer function and equivalent forms

If the input to an LTI system is $x(t)=\delta(t)$,

$$y(t)=h(t)*x(t)=h(t)*\delta(t), \tag{8}$$

where $y(t)$ is the output signal and $h(t)$ is the system's impulse response in the time domain. The Laplace transform of the input is $X(s)=1$, and the corresponding output yields

$$Y(s)=H(s)\cdot X(s)=H(s)\cdot 1. \tag{9}$$

Eq. 9 can be rewritten as

$$H(s)=\frac{Y(s)}{X(s)}. \tag{10}$$

The Laplace transform of $h(t)$ is $H(s)$, which is called the transfer function of the system with the complex frequency $s=\sigma+j\omega$.

Differential equation form of H(s)

$$\sum_{k=0}^{N} a_k \frac{d^k y(t)}{dt^k} = \sum_{k=0}^{L} b_k \frac{d^k x(t)}{dt^k}. \tag{11}$$

Using properties of the Laplace transform (linearity and derivative properties) yields

$$Y(s)\cdot\left[\sum_{k=0}^{N} a_k s^k\right]=X(s)\cdot\left[\sum_{k=0}^{L} b_k s^k\right]. \tag{12}$$

Eq. 10 can be found as the ratio of two polynomials,

$$H(s)=\frac{Y(s)}{X(s)}=\frac{\sum_{k=0}^{L} b_k s^k}{\sum_{k=0}^{N} a_k s^k}. \tag{13}$$

One can see that the input coefficients are related to the numerator terms in $H(s)$, and the output coefficients determines the denominator of $H(s)$.

Rational polynomial fraction (pole-zero) form of H(s)

$H(s)$ can be factored into first-order terms in s,

$$H(s)=\frac{K\prod_{i=1}^{L}(s-n_i)}{\prod_{k=1}^{N}(s-d_k)}, \tag{14}$$

where K is the scale factor. The s values, which make $H(s)$ zero ($s=n_i$) and infinity ($s=d_k$), are called the system's zero and pole, respectively.

Partial fractions with residual form of H(s)

Moreover, a partial fraction of $H(s)$ with a residual can be achieved by using the vector fitting procedure [8],

$$H(s)=\sum_{m=1}^{N}\frac{r_m}{s-p_m}+d+es, \tag{15}$$

where p_m is pole and r_m is residual; d and e are real values, whereas p_m and r_m can be either real or complex values. If e is zero, then the number of poles and zeros in the system is the same.

Stability of LTI system

The LTI system's stability can be discussed via the impulse response, the transfer function, and the system's poles and zeros. In terms of the region of convergence (ROC), the imaginary axis of the s-plane is included in ROC for a stable system. Specifically, for the system to be stable, all poles are in the left half plane (LHP) in a causal system case, whereas all poles must be in the right half plane (RHP) in an anti-causal system case (see the Table 1).

Pole-Zero Fitting Algorithm

A data fitting process is required to examine the experimental data in systematic ways. Gustavsen and Semlyen's [8] and Prony's [9] algorithms are employed in this study to analyze our physical system, the BAR. The comparison using two methods for our system is discussed here.

Gustavsen's method: rational approximation

Gustavsen and Semlyen [8] introduce the vector fitting

Table 1. Comparison of causal and anti-causal system stability

	Causal systems	Anti-causal systems
Impulse response	$h(t)=\sum_{k=1}^{N} C_k e^{d_k t}u(t)$	$h(t)=\sum_{k=1}^{N} C_k e^{d_k t}u(-t)$
Transfer function	$H(s)=\sum_{k=1}^{N}\dfrac{C_k}{s-d_k}$ $Re[s]>max\{Re[d_k]\}$	$H(s)=-\sum_{k=1}^{N}\dfrac{C_k}{s-d_k}$ $Re[s]>min\{Re[d_k]\}$

procedure. This method is also called pole-fitting or rational approximation. Assuming that we have data F(s), the aim is to find all R_i, Q_i, D, and E in a residual expansion of the form $\hat{F}(s)$ (model) in Eq. 16 by minimizing the (weighted) least-squares deviation from the measured data.

$$\hat{F}(s) = \sum_{i=1}^{N} \frac{R_i}{s - Q_i} + D + Es. \tag{16}$$

It is important to note that in our study, the data are only available as a function of ω; thus, we can rewrite the data F(s) as $F(\omega)$, which is related to the fitted function $\hat{F}(s)$ by $F(\omega) \approx \hat{F}(s)|_{s=j\omega}$. A two-step process is performed in this fitting procedure, pole identification and residual identification. This procedure is accomplished by converting a nonlinear least squares problem to a linear least squares problem by introducing an unknown scaling function $\sigma(s)$ and known poles (q_i in Eq. 17).

As an initial step, the explicit algebraic expression of $\sigma(s)$ and $\sigma \cdot \hat{F}(s)$ is defined by Gustavsen and Semlyen [8] using the vector system of equations,

$$\begin{bmatrix} (\sigma \cdot \hat{F}(s)) \\ \sigma(s) \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{N} \frac{p_i}{s - q_i} + d + es \\ \sum_{i=1}^{N} \frac{b_i}{s - q_i} + 1 \end{bmatrix}. \tag{17}$$

$(\sigma \cdot \hat{F}(s))$ and $\sigma(s)$ share the same known poles q_i and p_i, d, and e terms must be found. Multiplying data $F(\omega)$ with the scaling function $\sigma(s=j\omega)$ and evaluating over the many available frequency data points in $F(\omega)$, we can overdetermine this linear problem to estimate the unknown values of p_i, b_i, e, and d. The mathematical expression yields,

$$\sigma(j\omega) F(\omega) = (\sigma \hat{F})(j\omega) \rightarrow \frac{(\sigma \hat{F}(j\omega))}{\sigma(j\omega)} = \hat{F}(j\omega) \approx F(\omega). \tag{18}$$

Once we estimate the values of p_i, b_i, e, and d in Eq. 17, the equation can be redefined as a product form

$$\begin{bmatrix} (\sigma \cdot \hat{F}(s)) \\ \sigma(s) \end{bmatrix} = \begin{bmatrix} \dfrac{\tilde{e} \prod_{i=1}^{N+1} (s - \zeta_i)}{\prod_{i=1}^{N} (s - q_i)} \\ \dfrac{\prod_{i=1}^{N} (s - z_i)}{\prod_{i=1}^{N} (s - q_i)} \end{bmatrix}. \tag{19}$$

Therefore, our fitting function $\hat{F}(s)$ in Eq. 16 can be rewritten as

$$\hat{F}(s) = \sum_{i=1}^{N} \frac{R_i}{s - Q_i} + D + Es = \frac{\tilde{e} \prod_{i=1}^{N+1} (s - \zeta_i)}{\prod_{i=1}^{N} (s - z_i)}. \tag{20}$$

Thus, the zeros of the scaling function $\sigma(s)$ (z_i in Eq. 19) become the poles of the model function $\hat{F}(s)$ in Eq. 20; therefore, $z_i = Q_i$. Gustavsen and Semlyen [8] report that the residuals for $\hat{F}(s)$ (R_i, D, and E) can be calculated from p_i, d, e, and b_i by using z_i as the starting poles for $\hat{F}(s)$ (Q_i). Once this procedure is completed, we can approximate our data F(s) from the fitting function $\hat{F}(s)$, which yields $R_i = p_i$, D=d, and E=e (Eq. 16 and Eq. 17).

This algorithm is implemented by the method of Gustavsen and Semlyen [8], while applying for function rationalfit.m of the MATLAB (MathWorks, Inc., Washington DC, USA). This is a reasonable way to automate a fitting to within a given error tolerance. To facilitate an understanding of this algorithm, the procedure of this function is briefly explained as follows: Starting from the initial pole, the algorithm increases the order of the poles and searches for the best fit until the fit reaches the specified error tolerance level. An appropriate selection of starting poles is necessary for the convergence of the vector fitting method. For data with a resonance peak in magnitude (i.e., our data), Gustavsen and Semlyen [8] suggested that the complex conjugate pairs of the starting poles should be linearly distributed to achieve the best fitting result. Differences between the starting poles and the fitted poles can cause discrepancies between $\sigma(s)$ and $(\sigma \cdot \hat{F})(s)$ [8].

As the complex conjugate pairs of poles q_i (Eq. 19) order changes, they converge to the true poles. Then the scaling function $\sigma(s)$ will become 1, meaning that the calculated zeros z_i of $\sigma(s)$ have values arbitrarily close to the values of the starting poles, q_i. This single process can iterate over the data's frequency range until the fitting reaches the best fit. When it returns $\Re z_i > 0$ referring to an unstable pole in $\hat{F}(s)$, this program reflects the real parts in the LHP before the next iteration. Thus, the fitting result forces a system to be stable.

One can also define the e and d values as initial conditions in that they affect the total number of fitted zeros, namely, N_z. For example, N_z is the same as the number of poles (N) when e is zero but d is not zero. If e and d are zero, we can have one fewer zeros than the poles because the numerator order is one less than the denominator (see Eq. 17). Similarly, when d and e are both nonzero, N_z is N+1. In this study, we use the fitting condition with e zero and d nonzero, allowing us the same number of poles and zeros. We reach a slightly different fitting result by varying the initial value of e and d, but this does not critically affect the overall performance of the fitting result. Eventually, when this augmented equation is fitted to the given data $F(\omega)$ iterating toward $\sigma(s)=1$, the order of the poles for the fitting function meets the given error tolerance. The error is calculated using a mean squared error (MSE), in decibels, relative to the L2 norm of the signal:

$$MSE = 10\log_{10}\frac{\sum|F(s) - \hat{F}(s)|^2}{\sum|F(s)|^2} \text{ (dB)}. \tag{21}$$

Eq. 21 returns one single number (total error, not varying frequency) to evaluate goodness of fit. Note that in our study, we allow an MSE tolerance of -30 dB, which corresponds to a root mean square relative error of 3.16%.

Prony's method

Another popular data fitting procedure used in this study is called the Prony method (PM) (developed by de Prony in 1795) [9]. It models data as a linear combination of exponentials. However, the performance of the PM is known to be poor when the data contain noise [10]. This method returns a large biased result due to the noise; for example, it fits exponentials to any additive noise present in the signal so that the damping terms become much greater than the actual values [10]. To derive the mathematical formulation for the original Prony analysis, we can define a fitting function, $\hat{y}(n)$,

$$\hat{y}(n) = \sum_{i=1}^{L} A_i e^{\sigma_i t} \cos(2\pi f_i t + \varphi_i), \tag{22}$$

which is the sum of complex damped sinusoids, where A_i, σ_i, φ_i, and f_i are amplitude, damping coefficient, phase, and frequency of component i, respectively. Also, L is the total number of the damped exponential components. This function estimates given data $y(t)$, which consists of N samples $y(t_k) = y[k]$, where k=0, 1, 2, ..., N-1 are evenly spaced.

Via the Eulers theorem, the $\cos(2\pi f_i t + \varphi_i)$ term can be re-written as a sum of exponentials:

$$\cos(2\pi f_i t + \varphi_i) = \frac{1}{2}(e^{j2\pi f_i t}e^{j\varphi_i} + e^{-j2\pi f_i t}e^{-j\varphi_i}). \tag{23}$$

Plugging Eq. 23 into Eq. 22 and letting t=kT when T is the sampling rate,

$$\hat{y}(t) = \sum_{i=1}^{L} \frac{A_i}{2} e^{j\varphi_i} e^{(\sigma_i + j2\pi f_i)T} = \sum_{i=1}^{L} C_i \mu_i^k, \tag{24}$$

where $C_i = \frac{A_i}{2}e^{j\varphi_i}$ and $\mu_i = e^{(\sigma_i + 2j\pi f_i)T}$.

The PM searches for the C_i and μ_i in three steps: First, it solves the linear prediction model, which is constructed by the observed data set. Second, it finds roots of characteristic polynomials formed from the linear prediction coefficients. Third, it solves the original set of linear equations to obtain estimates of the exponential amplitude and sinusoidal phase.

Based on these concepts, in this study we have utilized MATLAB's built-in function prony.m to execute the Prony analysis. This function is for the time-domain design of infi-nite impulse response filters. It models a signal by employing a user-specified order of poles and zeros. In our study, we use the same numbers of poles and zeros (14 for each) taken from Gaustavsen's previous fitting results to compare the two fitting results. The PM assumes that frequencies are not periodic, which means they are not a sum of harmonically related sinusoidals. If we want to fit irrational data, we can use the Prony algorithm. For example, if we want to add two irrational sine waves [such as $\sin(\pi)$ and $\sin(\sqrt{\pi})$], then PM works. However, PM does not consider causality since there is no periodicity in the data. Therefore, this method is not relevant to our case as our data are causal supported through a Laplace transform. We compute the impulse response of the given frequency function $(H(\omega)^2)$ via the inverse fast Fourier transform. The impulse response shows that our data are periodic in time. Therefore, Gustavsen's method is appropriate in our case. As a result, we achieve a better data fit result from Gustavsen's method. Details of the result are discussed later in this document.

Application with Experiments

The electrical input impedance measurement method [3] that is used to calculate Hunt's parameters (Eqs. 2-5) is discussed here. Then we explain an explicit way to compute the transfer function from Hunt's parameters. The pole-zero fitting result of the transfer function is discussed at the end of this section.

Electrical input impedance measurements

The electrical input impedance measurement method and data have been adapted from Kim and Allen [3] to compute Hunt's parameters of the BAR as functions of the frequency. The chirp signal has been generated via MATLAB software on a laptop and sent to the BAR. The receiver has been maintained in the series position with a known value resistor such as 1 kΩ which source is grounded. Then the electrical input impedance Z_{in} has been computed based on Ohm's law. The signal-to-noise ratio is enhanced by averaging the response over more than 10 consecutive frames, and the first response block is dropped to consider the steady-state response.

In Kim and Allen [3] work, seven acoustic loads (different lengths of tubes including the no-load condition) are used for the measurement and seven corresponding electrical input impedance results are saved. The tube lengths are 0.2540, 0.3785, 0.8839, 1.2497, 2.3927, and 3.0658 cm. The same type tube is used with 0.15 cm of inner diameter, which is similar to the outer diameter of the BAR acoustic port (see the Fig. 7 of previous work from Kim and Allen [3]). Follow-

ing Weece and Allen [11], a mathematical formation of measured Z_{in} using Hunt parameters (Eqs. 2-5) is

$$Z_{in} \equiv \frac{E}{I} = Z_e + \frac{T_a^2}{Z_L + Z_a}, \tag{25}$$

where the acoustic load impedance Z_L is

$$Z_L \equiv -\frac{P}{U}. \tag{26}$$

For the blocked-end tube indicating U=0, $Z_L \approx Z_0$ coth (a·tube length) where Z_0 and a are the characteristic impedance of a tube and the complex propagation function, respectively. They are computed considering the viscous and thermal losses by Keefe [12] for calculation c=334.8 m/s, assuming 20°C room temperature.

Z_{in} is sensitive to the acoustic load conditions; thus, each condition of Z_{in} can be denoted as $Z_{in}|_A$, $Z_{in}|_B$, ..., $Z_{in}|_F$ excluding the no-tube condition. There are three unknowns (Hunt parameters) in Eq. 25; to compute one set of unknowns, three conditions of Z_{in} must be selected. The specific computation of the Hunt parameters using three input impedances is explained in Weece and Allen [11] and Kim and Allen [3].

Three cases of Z_{in} in the BAR are randomly chosen to calculate the Hunt parameters (please see a Fig. 11 of the published paper by Kim and Allen [3]). Hunt parameters explain the intrinsic characteristics of the system; thus, calculating Hunt parameters is one way to calibrate the system. By manipulating these data, the system's transfer function is computed as described in the next section.

Transfer function from Hunt's parameters

The transfer function, H(ω), of the system is calculated from Hunt's parameters and is compared with the direct pressure measurement data. There is reasonable agreement among these measures up to 6−7 kHz. The mathematical definition of H(ω) from Hunt's parameter is pressure over voltage (P/E) with a zero volume velocity (U=0),

$$H(\omega) = \frac{T_a}{Z_e}\bigg|_{U=0} = \frac{P}{E}\bigg|_{U=0}. \tag{27}$$

Note that this is the transduction impedance (one of Hunt's parameters), while Thevenin pressure over voltage in Eq. 27 is defined as the transfer function of this system. The dark-green line in Fig. 1 is taken from the pressure measurement using an ER-7C probe microphone (Etymotic Research, Inc., Elk Grove Village, IL, USA) as real-world reference data. To compute H(ω), the pressure data are divided by the input voltage (E_{in}) across the two electrical terminals of ED7045 (Knowles, Itasca, IL, USA) (U=0). Other than this direct pressure measurement, all responses are derived from the Hunt parameter calculation, using the electrical input impedance measurements data with various acoustical loads.

Results

The pole-zero fitting

The two vector fitting procedures are performed using Gustavsen's and Prony's methods. As discussed earlier, the former method provides a better data fitting result (shown in Fig. 2, 3). To compare the goodness of each method, we cal-
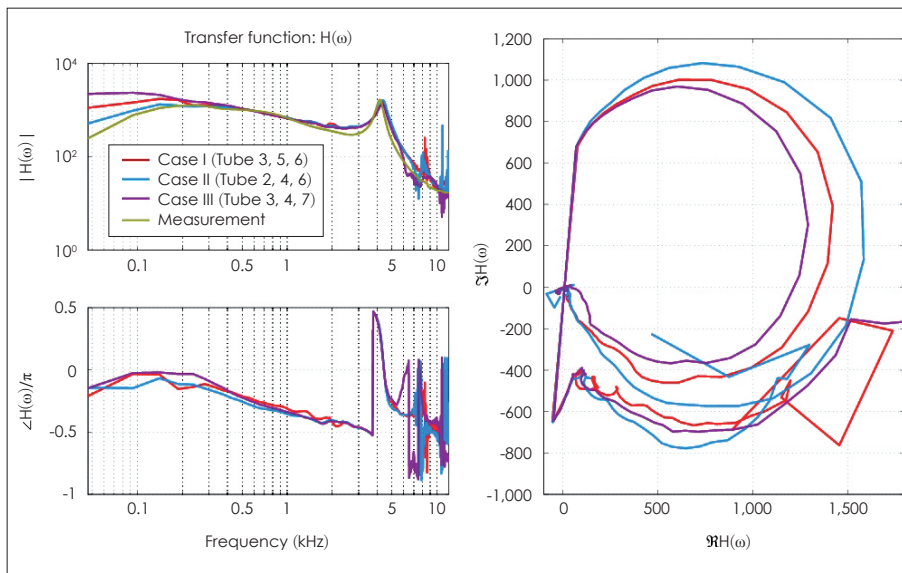

Fig. 1. Calculated transfer functions from various sources. The magnitude and phase data of H(ω) are shown in the left panel and the real and imaginary part of H(ω) are in the right panel. There are four different lines; the first three are calculated from the electrical experiments (Hunt parameters), and the last pressure data (in dark-green) are taken from the pressure measurement and are divided by the electrical input voltage. All data assume the blocked condition, U=0.

Fig. 2. Comparison of the data (black) and the fitting result (red) by Gustavsen and Semlyen (1999) [8]. A: Magnitude square of transfer function H. B: Phase square of transfer function H.
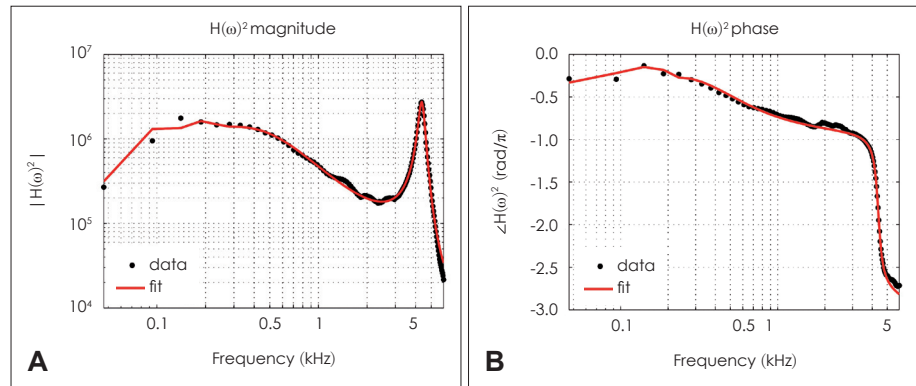

Fig. 3. Impulse response [IFFT of $H(\omega)^2$] (A) for the Prony method [9] and fitting estimation by the Prony method (B). IFFT: inverse fast Fourier transform.
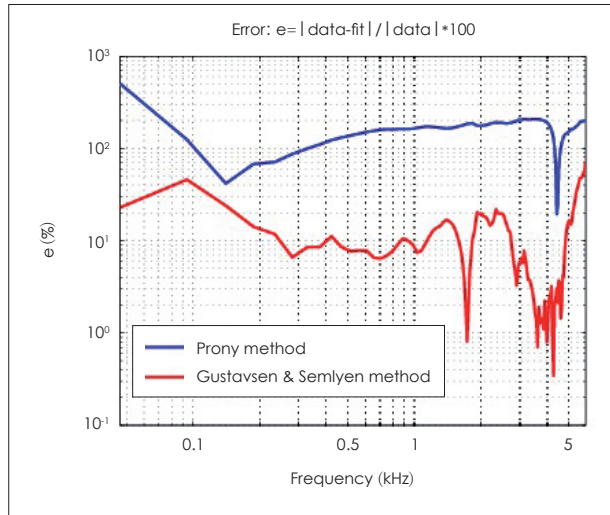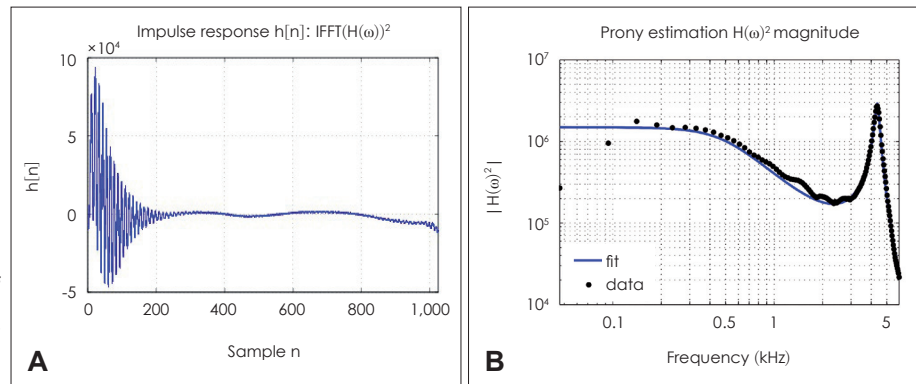

Fig. 4. Fitting error in %: e(f)=|data-fit|/|data| where f is a frequency. Gustavsen and Semlyen's [8] fitting result shows a lower error rate over the frequency region.

Fig. 5. Pole-zero plot of $H(\omega)^2$. Zeros are marked with blue O symbols. Poles are marked with red X symbols.

culate each error measured across frequencies,

$$e(f) = \frac{|\text{ data-fit }|}{|\text{ data }|},$$

(28)

where f is frequency. In Fig. 4, we can observe high error rates in low frequencies and minimum error rates at the peak ampli-
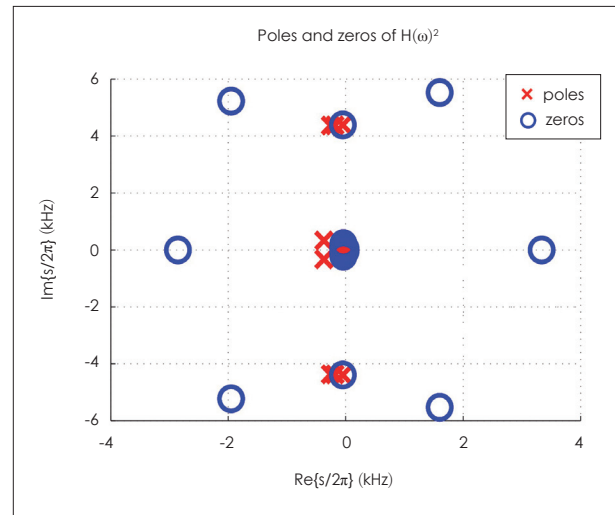
tude (resonance) of the data for both fitting methods. Overall lower error rates are observed in Gustavsen's fitting results. For this study, therefore, we choose the better result data for further analysis.

We limit the fitting frequency range up to 6 kHz, where $H(\omega)$ data start to show noise. The fitting result reflects reasonable agreement with the data within the frequency region
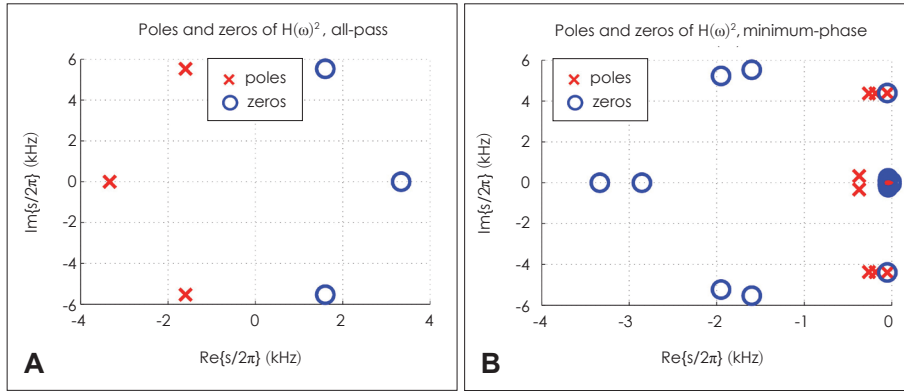
Fig. 6. Pole-zero plots for the all-pass (A) and the minimum-phase parts (B) of $H(\omega)^2$.

of our interest (Fig. 2).

Due to the π jump of $H(\omega)$'s phase in Fig. 1, the fitting results of $H(\omega)^2$ are much better than those of $H(\omega)$. Thus, $H(\omega)^2$ is used as our fitting data. The π jump happens during the Hunt's parameter calculation procedure in MATLAB programming when computing T_a from T^2. As a result, we can expect double poles and zeros for $H(\omega)^2$ compared to $H(\omega)$, but this does not affect the system analysis as the double poles and zeros are on top of each other. Fig. 5 shows the pole-zero plot of $H(\omega)^2$.

Poles and zeros on the $j\omega$ axis in Fig. 5 are on top of each other, meaning that they can cancel each other out. In this pole-zero fitting procedure, all poles are forced to be in LHP to keep the result stable; therefore, the region of convergence can be on the right side of the s-plane from the right-most pole location including the $j\omega$ axis.

All-pass/minimum-phase decomposition of H(ω)

A causal stable filter, $H(\omega)$, can be factored into a causal stable all-pass part cascading with a minimum-phase part,

$$H(\omega)=H_{AP}(\omega) \cdot H_{MP}(\omega), \qquad (29)$$

where $H_{AP}(\omega)$ and $H_{MP}(\omega)$ are an all-pass filter and a minimum-phase filter, respectively. Without changing the amplitude of the original filter, $H(\omega)$, the maximum-phase zeros in RHP of $H(\omega)$, are reflected in LHP to compose $H_{MP}(\omega)$, whereas $H_{AP}(\omega)$ has the original maximum-phase zero (Fig. 6).

A lossless all-pass filer, $H_{AP}(\omega)$, has a frequency response where the magnitude is always 1,

$$H_{AP}(\omega)=| H_{AP}(\omega) | \, e^{j\angle H_{AP}(\omega)}=1 \cdot e^{j\varphi(\omega)}, \qquad (30)$$

where $H_{AP}(\omega)=\varphi(\omega)$. Based on Fig. 7, we can approximate $\varphi(\omega)=\omega T$, yielding

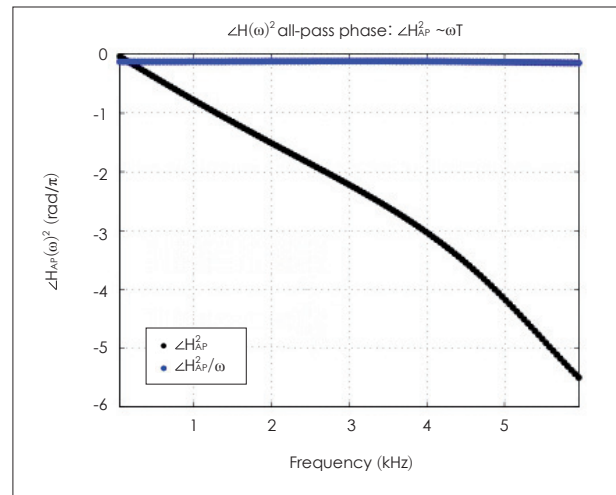$$H_{AP}(\omega)^2=e^{j\omega 2T}=e^{j2\varphi(\omega)}, \qquad (31)$$

Fig. 7. Phase of $H_{AP}(\omega)^2$. It can be approximated as a linear phase filter.

where $2T=2\varphi(\omega)/\omega$ is a constant.

The group delay $\tau_g(\omega)$ of $H_{AP}(\omega)^2$ can be calculated as follows:

$$\tau_g(\omega)=-\frac{2d\varphi(\omega)}{d\omega}. \qquad (32)$$

The result is shown in Fig. 8.

If one part of an entire system is approximated to be lossless, then this all-pass part can be used to represent the lossless part of the system. The BAR system is composed of three sections: electrical, mechanical, and acoustic. An electrical signal goes into the electrical terminal of BAR, passes through the mechanical part (magnets), and then is coupled to the acoustic part (diaphragm) to create sound (Fig. 9).

Assuming that the acoustic part of BAR is a lossless system, we can relate $\tau_g(\omega)$ to an acoustic delay of the sound. Considering the squaring factor, the acoustic group delay is approximately 50 μs below 3.5 kHz (Fig. 8).
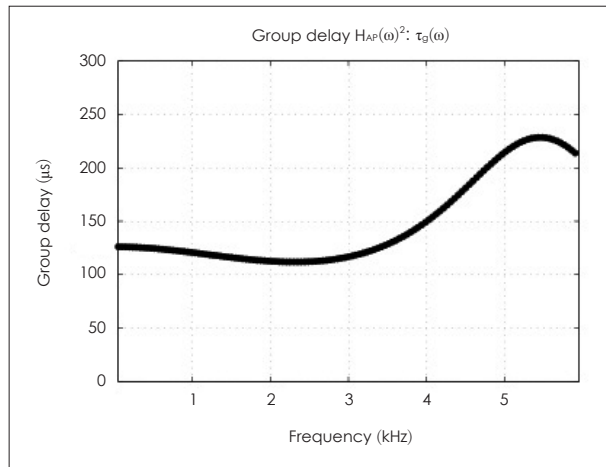
Fig. 8. The group delay (τ_g) of $H_{AP}(\omega)^2$ as a function of frequency.
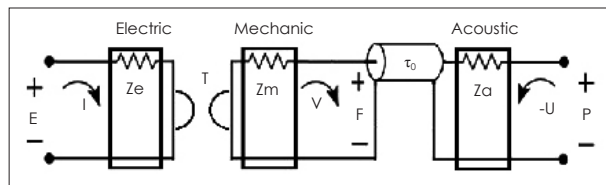

Fig. 9. A simple schematic diagram of a balanced armature receiver system based on energy transformation flow, from an electrical input to an acoustic output. A gyrator is coupling between the electrical part and mechanical part via transduction coefficient T. A delay τ is represented by a transmission line referring to a pure acoustic delay from mechanic to acoustics.

Conclusions

Beginning with a review of basic DSP concepts such as LTI system, transfer function, and pole-zero fitting, we have analyzed the BAR system by performing vector fitting of its transfer function H(ω). One interesting point is that H(ω) of this electro-acoustic system is computed solely from the electrical side by calculating Hunt's parameters from the electrical input impedance data, which are varying acoustic loads. This electrical impedance variation proves the existence of the motional impedance due to acoustic load. This method carries a relatively low cost for computing a transfer function of a hybrid system such as an electro-acoustic or electro-mechanic transducer. Moreover, we can assume that the BAR system is a stable, causal LTI system from the pole-zero fitting of H(ω) result using Gustavsen and Semlyen's method [8]. By factoring out the all-pass part of H(ω) and examining its group-delay, the system's delay is estimated.

Prony's fitting method [9] is also reviewed and we conclude that this method is inappropriate for our causal data. However, only a few cases (in terms of either H(ω) or type of a transducer) have been included in the current study. For further investigation, we should extend the present study to compare and analyze H(ω) of several different BAR types. Also, characterizing the physical and mathematical properties of BAR via the pole and zero locations may be another interesting topic that provides clinical implications.

Conflicts of interest

The authors have no financial conflicts of interest.

REFERENCES

1) Robinson SR, Nguyen CT, Allen JB. Characterizing the ear canal acoustic impedance and reflectance by pole-zero fitting. Hear Res 2013;301:168-82.
2) Hunt FV. Electroacoustics: the analysis of transduction, and its historical background. Harvard University Press;1954.
3) Kim N, Allen JB. Two-port network analysis and modeling of a balanced armature receiver. Hear Res 2013;301:156-67.
4) Wegel RL. Theory of magneto-mechanical systems as applied to telephone receivers and similar structures. J Am Inst Elect Eng 1921; 40:791-802.
5) Van Valkenburg ME. Network analysis. 2nd ed. Englewood Cliffs, NJ: Prentice-Hall;1964.
6) Carlin HJ, Giordano AB. Network theory: an introduction to reciprocal and nonreciprocal circuits. 1st ed. Englewood Cliffs, NJ: Prentice-Hall;1964.
7) McMillan EM. Violation of the reciprocity theorem in linear passive electromechanical systems. J Acoust Soc Am 1946;18:344-7.
8) Gustavsen B, Semlyen A. Rational approximation of frequency domain responses by vector fitting. IEEE Trans Power Delivery 1999; 14:1052-61.
9) de Prony BR. Essai experimental et Analytique sur les lois de la dilatabilité des fluides élastiques et sur celles de la force expansive de la vapeur de l'eau et de la vapeur de l'alcool, à différentes temperatures. J l'Ecole Polytech 1975;1:24-76.
10) Marple SL. Digital spectral analysis with applications. Prentice-Hall; 1987.
11) Weece R, Allen J. A method for calibration of bone driver transducers to measure the mastoid impedance. Hear Res 2010;263:216-23.
12) Keefe DH. Acoustical wave propagation in cylindrical ducts: transmission line parameter approximations for isothermal and nonisothermal boundary conditions. J Acous Soc Am 1984;75:58-62.